AMD "*Pacifica*" Virtualization Technology

March 30, 2005

Virtualization

is the pooling and abstraction of resources

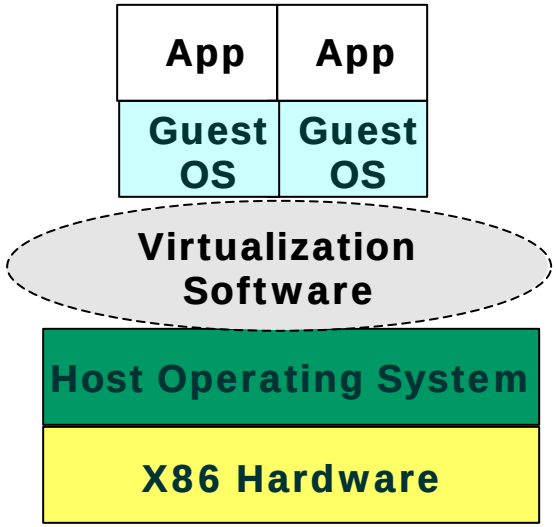
in a way that masks the physical nature and boundaries of those resources

from the resource users

AMD

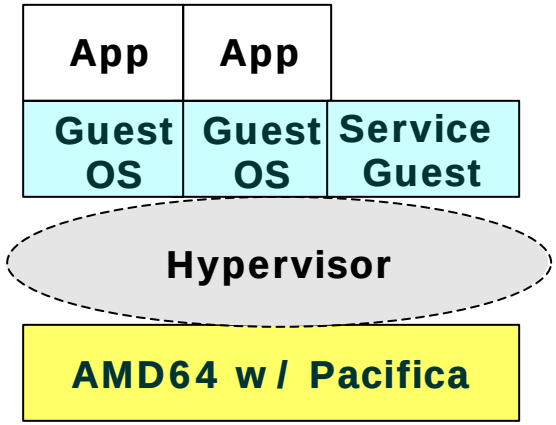
Carve a Server into Many Virtual Machines

Hosted Virtualization



- Virtualization software manages resources between Host and Guest OS's
- Application can suffer decreased performance due to added overhead

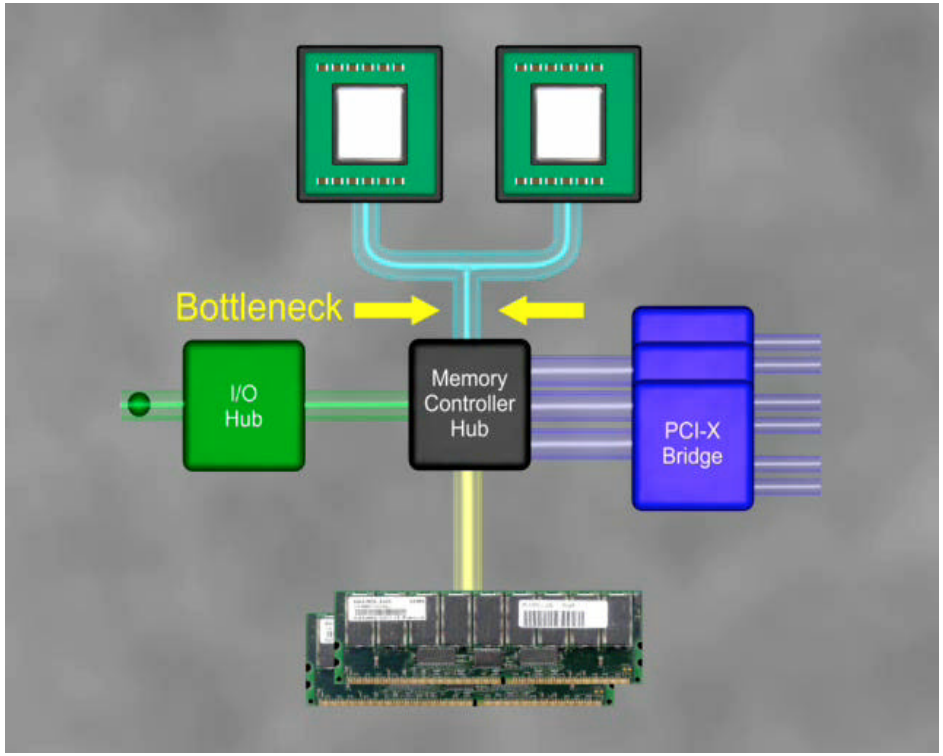
Hypervisor-based Virtualization



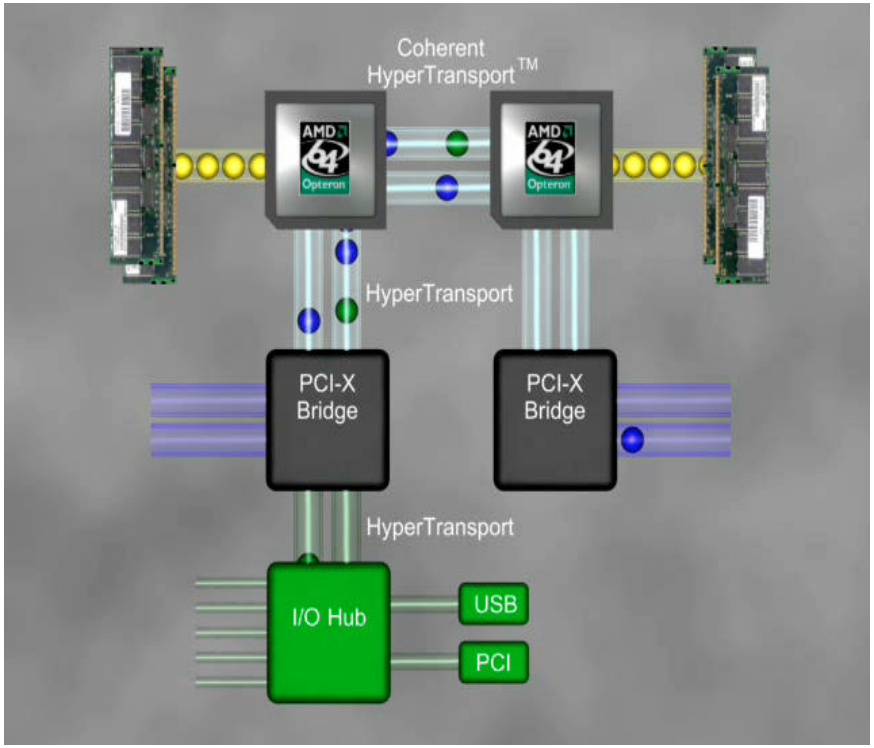
- Virtualization Software (Hypervisor) is the host environment.
- Enables better software performance by eliminating some of the associated overhead
- If Hardware is available, the Hypervisor can be designed to take advantage of it

System Architecture Makes a Difference

- Legacy Architectures based around front-side bus aren't scalable for today's virtualization needs



- AMD's Direct Connect Architecture reduces the bottlenecks, enabling efficient partitioning



Driving virtualization into the processor ***with Pacifica***!

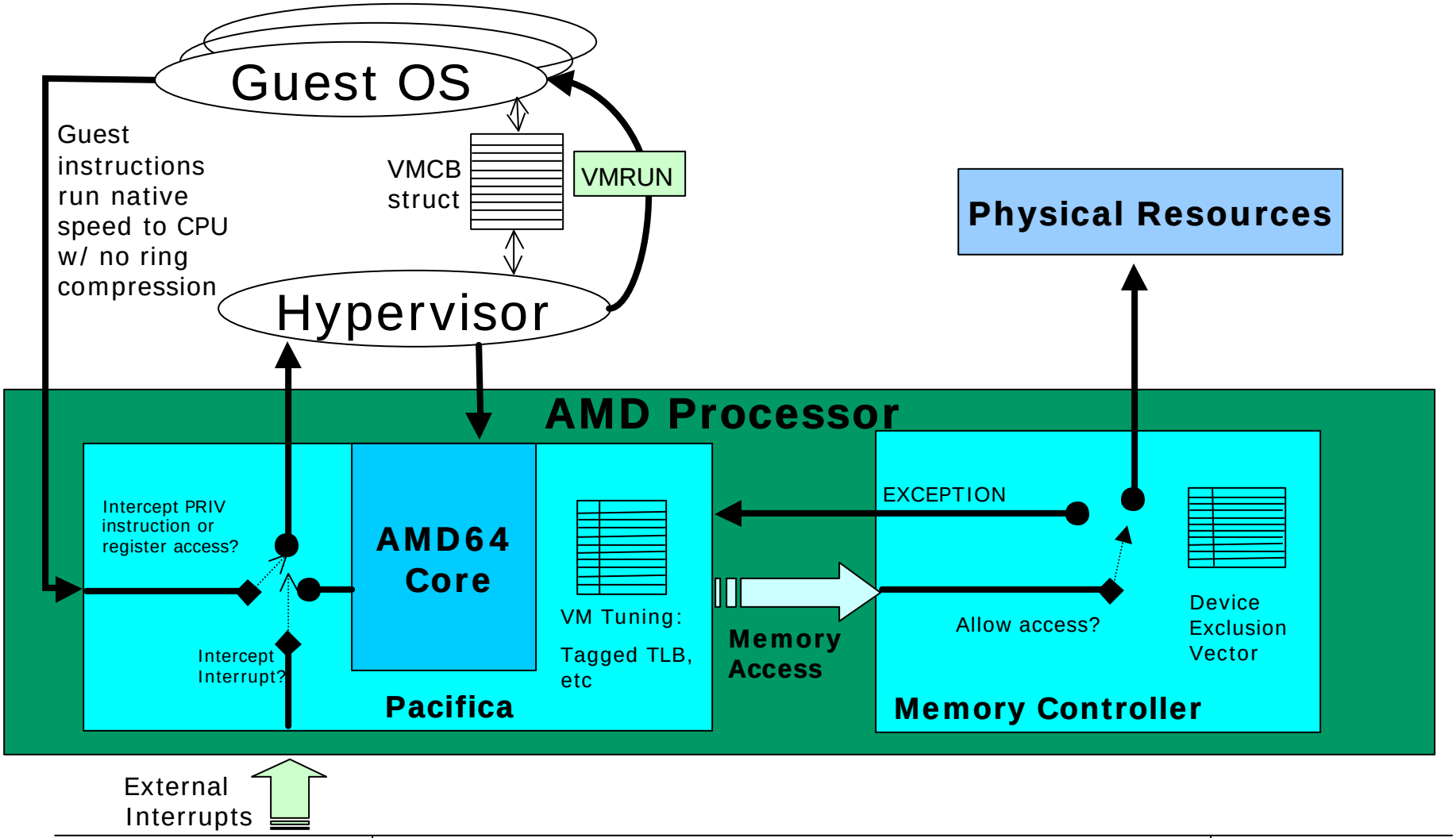
- Native virtualization of x86 architecture requires “unnatural acts” to achieve – leading to increased performance overhead, lower security, and increased complexity
- Moving functionality traditionally served by software-based hypervisor into the processor helps to solve these problems.
- ***PACIFICA is next logical evolution to the AMD’s Direct Connect Architecture* to provide technology for silicon enhanced virtualization**
- PACIFICA allows the software vendors to focus on the value-add, leaving the worry of proper emulation to the processor.

Pacifica virtualization technology allows AMD to continue to offer a competitive performance roadmap while meeting the system architecture demands of our customers

Pacifica Overview & Highlights

- Pacifica drastically reducing the complexity and performance impact of existing x86/64 virtualization
- **Pacifica** enabled parts will launch in AMD processors beginning in 1H'2006 across segments; mobile, server/workstation, and desktop markets
- Compatible with x86 and AMD64 applications – no change in legacy software is required.
- Virtualization and partitioned applications will experience the greatest **performance advantage**.
- AMD Opteron with Pacifica *enhanced virtualization* is a continuing example of how AMD is extending it's **Direct Connect Architecture** and **multi-core technology** leadership

Pacifica Silicon Enhanced Virtualization

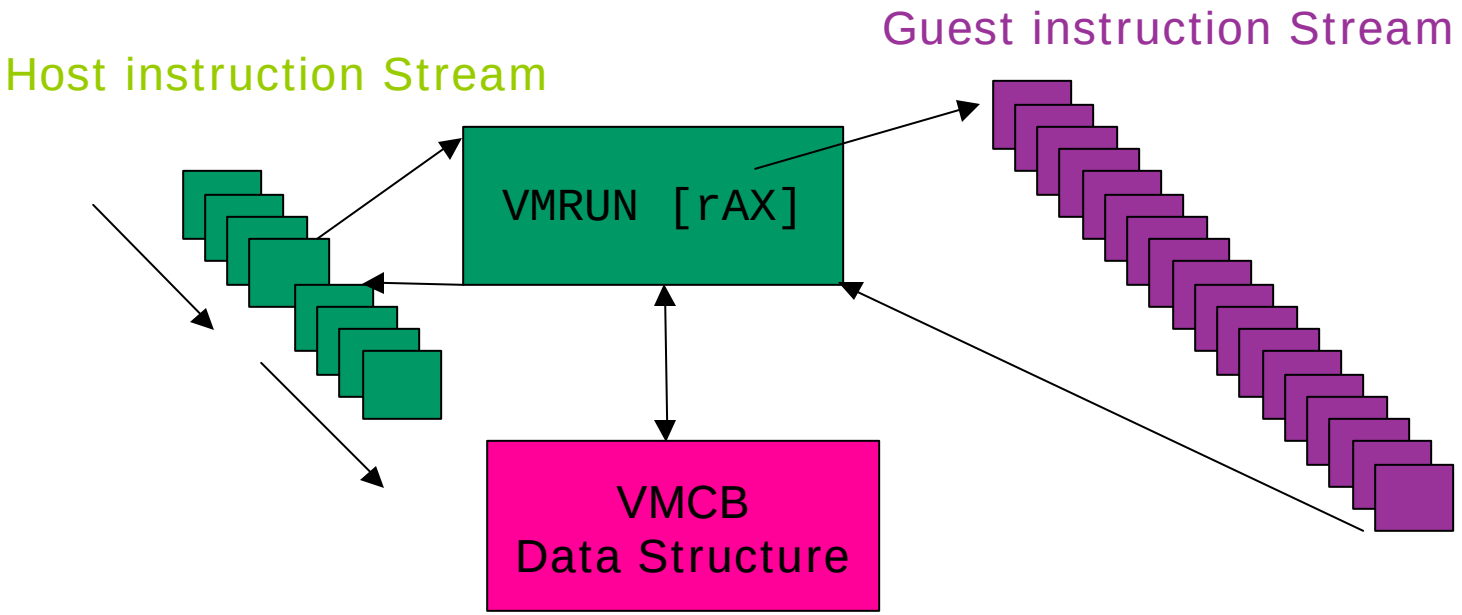


Pacifica Features to Accelerate & Secure Virtualization

- New Processor Mode: ***Guest Mode***
- New Data Structure: ***Virtual Machine Control Block (VMCB)***
- New Instruction: ***VMRUN***
- New memory mode: ***Real Mode w/ Paging***
- External Access Protection through ***Device Exclusion Vectors (DEV)***
- ***Selective Interception***, increasing performance and enabling para-virtualization
- Support for ***SKINIT*** ("secure kernel" init)
- ***Tagged TLB***
- ***Nested Page Table Support***
- ***Interrupt architecture changes***
- ***All instructions now Restartable***

Core Pacifica Architecture: VMRUN

- Virtualization based on Virtual Machine Run (VMRUN) instruction
- VMRUN executed by host causes the guest to run
- Guest runs until it exits back to the host
- World-switch: host → guest → host
- Host resumes at the instruction following VMRUN



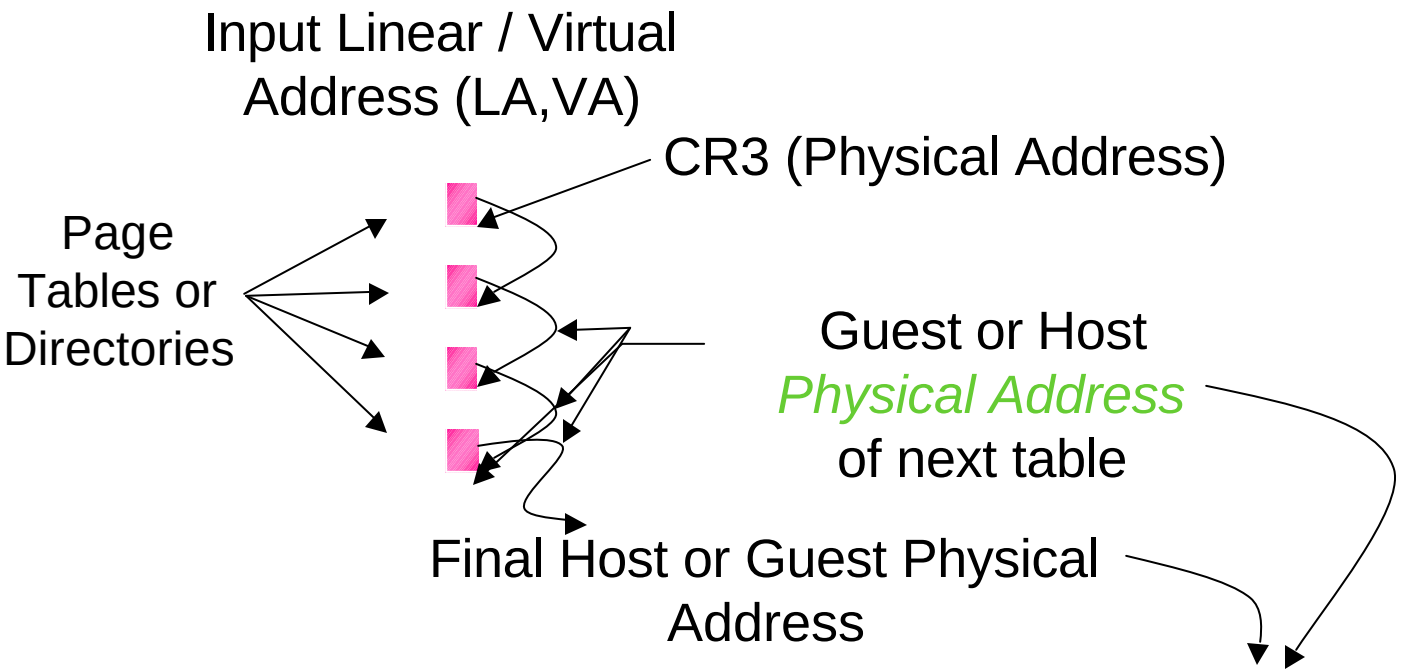
Core Pacifica Architecture: Intercepts

- Guest runs until:
  - It performs an action that causes an exit to the host
  - It explicitly executes the VMMCALL instruction
- The VMCB for a guest has settings that determine what actions cause the guest to exit to host
  - These intercepts can vary from guest to guest
  - Two kinds of intercepts
    - Exception & Interrupt Intercepts
    - Instruction Intercepts
  - Rich set of intercepts allow the host to set customize each guest's privileges
- Information about the intercepted event is put into the VMCB on exit

Core Pacifica Architecture: VMCB

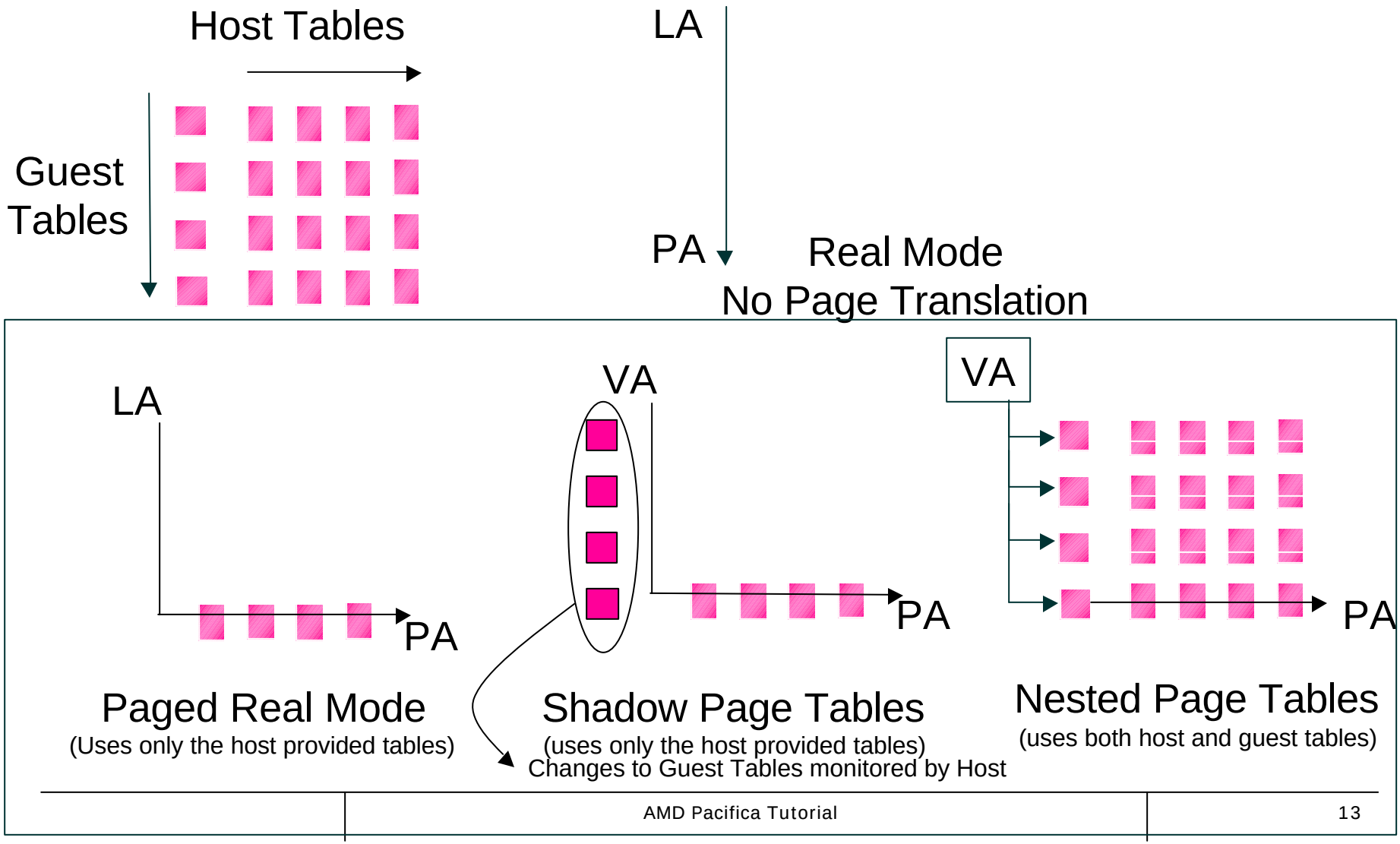
- All CPU state for guest is located in the Virtual Memory Control Block (VMCB)data structure
- VMRUN: Entry
  - Host state is saved to memory
  - Guest state loaded from VMCB
  - Guest runs
- VMRUN: Exit
  - Guest state is saved back to VMCB
  - Host state loaded from memory
- Host state saved using Model Specific Register (MSR): vm_hsave_pa

Address Translation: Page Tables



If this is a “Guest Physical” it must be translated to “Host Physical” via the host page tables when nested paging Is enabled.

Address Translation: Modes w/Virtualization



Core Pacifica Architecture: Shadow Page Tables

- Memory Protection – CPU accesses
  - Shadow Page Tables (SPT)
  - Nested Page Tables
- SPT Constraints on host design
  - Host intercepts guest CR3 Reads/Writes
  - Host monitors guest edits to guest page tables
    - Guest page tables are marked “read only”
  - Host constructs and manages SPT in software
    - Software strategies for this are mature
- Guest never sees the “real” page tables or the real content of Control Register 3 (CR3)
- Address Space ID’s (ASID) implemented to improve Translation Look-aside Buffer (TLB) performance
  - VMRUN sets guest ASID

Core Pacifica Architecture:
CPU Access Protection

- SPT sets guest access rights to physical address space
  - No guest access is possible unless a mapping is present in the SPT
  - Covers DRAM and Memory Mapped Input/Output (MMIO)
  - Minimum granularity 4k-bytes
- VMCB contains a pointer to an IO Permission Map (IOPM) that controls guest access rights to IO Ports
  - Granularity is to 1-byte port
- VMCB contains a pointer to an MSR permission map that control guest access to MSRs

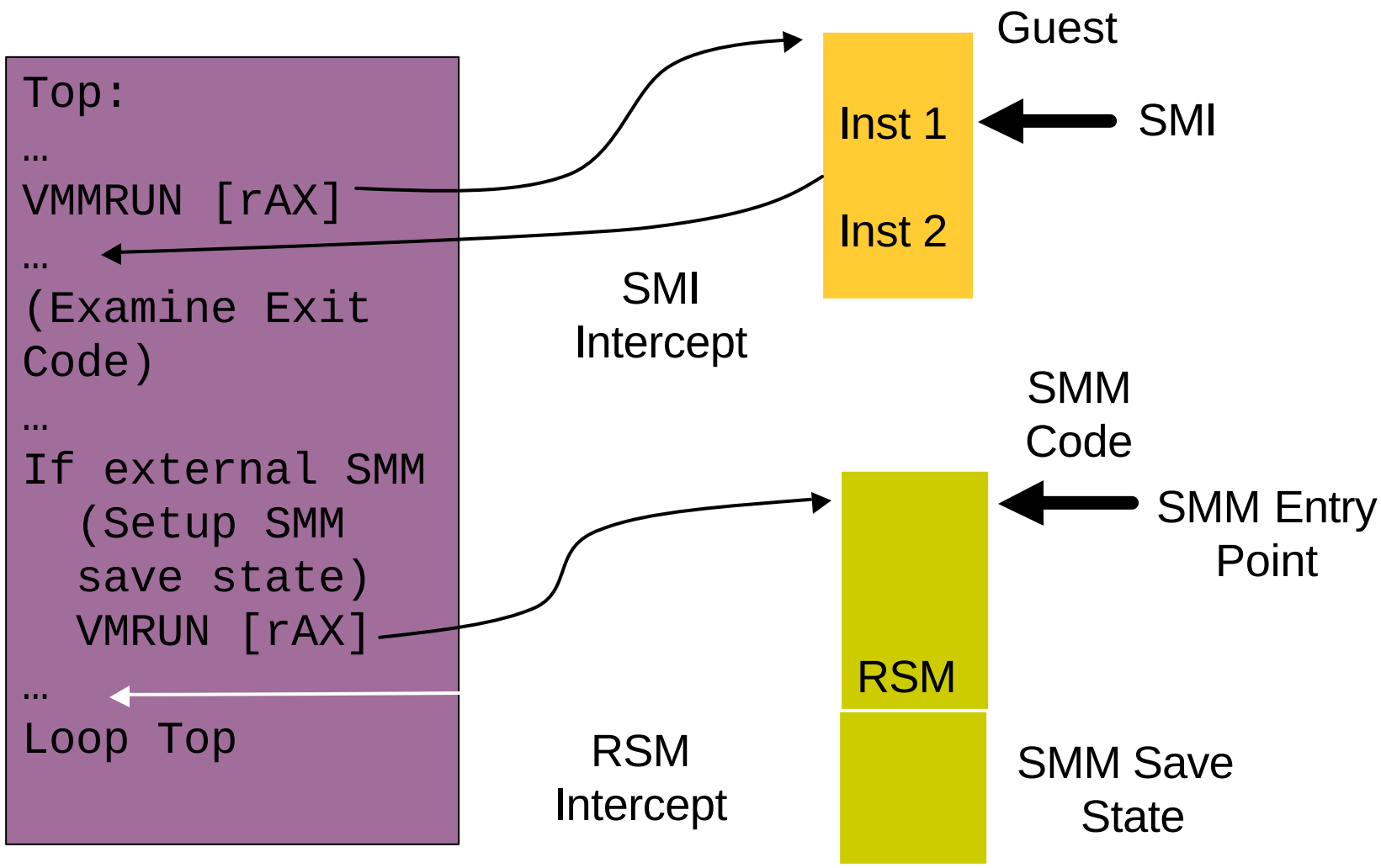
Core Pacifica Architecture: Interrupts

- Processor response to HW interrupts is setup in the VMCB
- Two Options:
  - Hardware interrupts while guest is running are intercepted causing exit to host
    - Host manages physical APIC
    - Host determines interrupt routing and distribution
    - Host injects virtual interrupts into guests as needed
    - Hardware support for virtual interrupts: v_irq, v_vector, v_prio , v_tpr, PHYS_IF
  - Interrupts serviced directly in the guest
    - Guest manages physical APIC
    - Host can still inject virtual interrupts
  - Global Interrupt Flag (GIF)
    - Protects host code critical-regions

Core Pacifica Architecture: SMM

- Pacifica implements a flexible architecture for System Management Interrupt (SMI)/SMM
  - Full legacy support for SMI from within host or guest
  - SMI Intercepts:
    - Allow host to scrub state if needed followed by native SMI from host
    - Support for “containerized” SMM
  - SMM Mode control via SMM_CTL_MSR
    - Allow host to scrub state and dispatch the SMM handler from a VMCB

Pacifica: Containerized SMM Flow



Pacifica: Paged Real Mode (New)

- SMM code is designed to start in real mode
- Memory protections rely on paging, guests *must* run with paging enabled
- Pacifica Solution: Paged Real Mode
  - Only available for guests
  - cr0.pg=1, cr0.pe=0
  - Host must intercept page faults
  - Real-mode address translation (segment+offset) = Linear address ➔ translation via SPT ➔ physical address
  - Correct composition of SPT's is host responsibility
    - Guest is assuming linear, 0-based mapping

Pacifica: DMA Protection

- Protection Domains
  - Mapping from bus/device ID to protection domain

- Device Exclusion Vector (DEV)
  - One DEV per protection domain
  - Permission-checks all upstream accesses
  - 1 bit per physical 4K page (0.003% tax; 128K / 4G) of the system address space
    - Protection for both DRAM and Memory Mapped IO space
  - *Contiguous* table in physical memory

Summary

- Virtualization is being used in several server scenarios today
- AMD expects that virtualization will prove valuable for PC clients too
- There are ways to modify the X86 architecture, so that virtualization is easier to accomplish, performs better, and provides more security
- AMD's *Pacifica* technology is being developed for future AMD64 CPUs for servers and clients
- Key technologies include adding new instructions, supporting different methods of handling page tables, handle host and guest interrupts (including SMI/SMM), and provide DMA protection